USING "GLIM" FOR COMPUTING THE RASCH MODEL AND THE CORRESPONDING MULTIPLICATIVE POISSON MODEL

R. DITTRICH*)
R. HATZINGER*)

Forschungsbericht/
Research Memorandum No. 167

November 1981

*) Scholaren der Abteilung für Mathematische Methoden und Computerverfahren am Institut für Höhere Studien, Wien

Abstract

Randomising the row marginals of Rasch's dichotomous logistic model (RM) leads to an extended randomised version of the RM under assumption of free variability of the marginal probabilities on the surface of a corresponding probability simplex. The resulting model is shown to be widely identical to a multiplicative poisson model. The part of the ML-equation for estimating the item parameters coincides with the conditional maximum likelihood of the RM. The connection of these models is formulated in the context of generalised linear models which provides the possibility to use the GLIM-package for computing numerical solutions for a moderate number of items. The general structure of the models is discussed by an example. Additionally the paper contains the presentation of GLIM-macros and directives required for individual applications.

Zusammenfassung

Randomisieren der Randsummen im dichotomen logistischen Modell von Rasch (RM) führt zu einer erweiterten randomisierten Version des RM, bei der die Wahrscheinlichkeiten für die Randsummen frei an der Oberfläche des entsprechenden Wahrscheinlichkeitssimplex variieren. Es wird gezeigt, daß dieses Modell im wesentlichen einem multiplikativen Poisson-Modell entspricht. Der Teil der ML-Gleichung zur Schätzung des Item-Parameters ist identisch mit der conditional maximum Likelihood des RM. Die Formulierung des Zusammenhangs zwischen diesen Modellen im Rahmen des verallgemeinerten linearen Ansatzes ermöglicht die Verwendung des Programmsystems GLIM zur Berechung numerischer Lösungen unter der Voraussetzung einer nicht zu großen Anzahl von Items. Die generelle Struktur der Modelle wird anhand eines Beispiels diskutiert. Zusätzlich enthält die Arbeit die zur allgemeinen Anwendung benötigten GLIM-macros und Steuerbefehle.

The authors are indepted to Dr. J.A. Nelder and Mag. M. Hudec for reading the manuscript and giving essential comments.

Vienna, November 1981

Contents

1. THE RASCH MODEL

Given a sample of subjects i, i=1,..,n, who have responded to objects (items) j, j=1,..,k, so that the observations y_{ij} can be regarded as realisations of a Bernoulli variable Y (coded by 1 or 0 respectively), RASCH (1960) suggested the model

$$(1.1) \quad \text{logit } p(y_{ij}=1) = \ln\{p_{ij}/(1-p_{ij})\} = \ln\xi_i + \ln\lambda_j$$

or

$$(1.2) \quad p(y_{ij}=1) = \frac{\exp(\xi_i+\lambda_j)}{1+\exp(\xi_i+\lambda_j)}$$

which is typically applied to item analysis taking individual differences into account. λ_j represents the "difficulty" of item j and ξ_i the "ability" of the responding person i. The observations provide a (nxk)-matrix $((y_{ij}))=Y$ with the marginals $r_i=\Sigma_j y_{ij}$ and $s_j=\Sigma_i y_{ij}$.
Considering (1.1) as a Generalised Linear Model (GLM) with binomial error and logit link

$$(1.3) \quad \text{logit } \mu = \eta = \ln(X\beta) = \ln\xi_i + \ln\lambda_j = p_{ij}$$

(for more detailled presentation of GLMs in the present context see Section 5.) the analogy to a binary regression of COX-type is obvious, when the data matrix Y is regarded as a (nxk)-vector ordered with respect to the covariates.

However, analysing given data in this way could in particular cases lead to some difficulties. Using the so-called unconditional maximum likelihood method (UML) for estimation of the parameters (equating the marginals r_i and s_j with their respective expectancies), ANDERSEN (1973) showed that the solutions for $\hat{\xi}$ and $\hat{\lambda}$ have an approximate asymptotic bias of k/(k-1) for moderate k. Additionally regarding the person parameters as nuisance the number of parameters tend to infinity when $n \to \infty$.

RASCH's proposal for avoiding these problems is based on the idea of conditioning on the row sums r_i, which leads to a conditional model where the person parameters do not occur in the ML-equations, as can be seen from the following arguments. Denoting A_i as a subset of items "correctly" responded to by person i, the probability for observing exactly that pattern of responses A is

$$(1.4) \qquad P(A_i=A) = \Pi_{j\varepsilon A}\, p_{ij}\, \Pi_{j\varepsilon A^c}\,(1-p_{ij}) =$$

$$\Pi_{j\varepsilon A} \frac{\xi_i\lambda_j}{1+\xi_i\lambda_j}\; \Pi_{j\varepsilon A^c} \frac{1}{1+\xi_i\lambda_j} =$$

$$\frac{\xi_i^{\,r}\; \Pi_{j\varepsilon A}\,\lambda_j}{\Pi_{j=1}^{k}\,(1+\xi_i\lambda_j)}$$

where $r = \#A$, which is the number of elements in A.

Let $\Sigma_{B:=B=r} P(A_i=B)$ be the sum of probabilities so that the response pattern A_i of person i equals =B, defined as the subset of all patterns of items with exactly r elements, the probability for the event $A_i=A$ conditioning on r_i is

$$(1.5) \qquad P(A_i=A\,|\,r_i=r) = \frac{P(A_i=A)}{\Sigma_{B:\#B=r} P(A_i=B)} =$$

$$\frac{\xi_i^{\,r}\,(\Pi_{j\varepsilon A}\lambda_j) \;/\; \Pi_{j=1}^{k}(1+\xi_i\lambda_j)}{\Sigma_{B:\#B=r}\,\xi_i^{\,r}\,(\Pi_{j\varepsilon B}\lambda_j) \;/\; \Pi_{j=1}^{k}(1+\xi_i\lambda_j)} =$$

$$\frac{\Pi_{j\varepsilon A}\lambda_j}{\Sigma_{B:\#B=r}\Pi_{j\varepsilon B}\lambda_j} = \frac{\Pi_{j\varepsilon A}\lambda_j}{\gamma_r(\lambda_1,\ldots,\lambda_k)};$$

where

$$\gamma_r(\lambda_1,\ldots,\lambda_k) = \Sigma_{B:\#B=r}\Pi_{j\varepsilon B}\ \lambda_j$$

is the "elementary symetric function"of order r.
Multiplication of (1.5) over i provides the conditional likelihood function L_c given the person's marginals r

$$(1.6)\qquad L_c(\lambda_1,\ldots,\lambda_k) = \Pi_{i=1}^{n} \frac{\Pi_{j\varepsilon A_i}\ \lambda_j}{\gamma_{r_i}(\lambda_1,\ldots,\lambda_k)} =$$

$$\frac{\lambda_1^{s_1}\lambda_2^{s_2}\ldots\lambda_k^{s_k}}{\Pi_{r=0}^{k}\ \gamma_r(\lambda_1,\ldots,\lambda_k)^{n_r}}$$

implying the possibility of estimating the item parameters λ_j independently of the person parameters. Excluding the trivial scoregroup r=0 and r=k the estimates obtained by maximizing L_c are consistent as $n \rightarrow \infty$. (For further discussions of consistency and asymptotic normality see ANDERSEN,1973,1980. FISCHER(1981) describes the conditions for existence and uniqueness of a conditional ML-solution) A Fortran-program, also used for parameter estimation in the present paper's example, can be found in FISCHER (1974).

As TJUR (1980) suggests it seems permissible to introduce an underlying common distribution of the person parameters, since in many applications of the RM it is reasonable to think of a set of subjects as a random sample of a larger population. Thus an underlying, unknown distribution π of the person parameters can be introduced by assuming the ξ_i's as independently, identically distributed random variables. The probability for observing $r_i = r$ is then given by

$$(1.7)\qquad q_r = q_r(\lambda_1,\ldots,\lambda_k,\pi) = \Sigma_{A:\#A=r} P(A_i=A) =$$

$$E(\Sigma_{A:\#A=r} P(A_i=A|\xi_i)) =$$

$$\int \Sigma_{A:\#A=r} \frac{\xi^r \Pi_{j\varepsilon A}\lambda_j}{\Pi_j(1+\xi\lambda_j)} \pi(d\xi) =$$

$$\gamma_r(\lambda_1,\ldots,\lambda_k) \int \frac{\xi^r}{\Pi_j(1+\xi\lambda_j)} \pi(d\xi)$$

and

$$(1.8)\qquad P(A_i=A) = q_r P(A_i=A|r_i=r)$$

as the probability for any response pattern with $r=\#A$. In analogy with the original model the conditional probability $P(A_i=A|r_i=r)$ is independent of ξ_i. L_r denotes the likelihood function for this randomised version of the RM, which is

$$(1.9)\qquad L_r(\lambda_1,\ldots,\lambda_k,\pi) = \Pi_{i=1}^n \frac{q_{r_i}\,\Pi_{j\varepsilon A}\,\lambda_j}{\gamma_{r_i}(\lambda_1,\ldots,\lambda_k)} =$$

$$\frac{\lambda_1^{s_1}\lambda_2^{s_2}\ldots\lambda_k^{s_k}}{\Pi_{r=0}^k \gamma_r(\lambda_1,\ldots,\lambda_k)^{n_r}}\; q_0^{n_0} q_1^{n_1}\ldots q_k^{n_k}$$

Since there exists no test to decide whether the marginals

r_i are random or fixed it seems a matter of interpretation which model is assumed to adequately describe the experiment. As TJUR states two different ways can be used to introduce random row effects. One is to assume an unknown distribution which determines the variation of the row parameters ξ_i, as formulated in the "random model". The other possibility can be described in the following way: First observing the row sums r_i assuming them as realisations of a random variable with distribution $q_r=P(r_i=r)$. Then the remaining variation can be assumed due to conditional distributions given the marginals r_i, which are independent of the λ_j´s. This means the observations y_{ij} are generated in accordance to the conditional distribution of the RM.

This second model is an extension to the randomised version of the RM. (accordingly TJUR calles it "extended random model") by assuming free variability of the vector $(q_0, q_1, \ldots, q_k)$ on the surface of a k-dimensional probability simplex with $\Sigma_{r=0}^{k} q_r = 1$, where $q_i \geq 0$ and $\lambda_j > 0$.
The Likelihood L_e for the extended model equals the likelihood L_r (1.9) but is regarded as a function of the λ_j´s and the q_r´s

(1.10) $$L_e(\lambda_1, \ldots, \lambda_k, q_0, q_1, \ldots, q_k) = L_c(\lambda_1, \ldots, \lambda_k)\; q_0^{n_0} q_1^{n_1} \ldots q_k^{n_k}$$

This model has previously been discussed by MARTIN-LÖF (1970), who showed the sufficiencyof the column sums s_j and the empirical distribution of the row sums n_r, $r=0, \ldots, k$.

2. THE CORRESPONDING MULTIPLICATIVE POISSON MODEL (MPM)

TJUR (1980) suggests another way of representing the observations y_{ij}, regarding the items j, j=1,..k, as k classifying factors. According to the person's responses ("correct" or "wrong") a 2^k-contingency table is obtained with cell frequencies n_A, where n denotes the number of subjects with response pattern A. This leads to a MPM with

$$(2.1) \qquad E(n_A) = \mu \, \Pi_{j \epsilon A} \, \lambda_j$$

which corresponds to an extended random model (1.10) assuming a common person parameter $\xi = \xi_1 = \ldots = \xi_n$. To take individual differences into account an additional factor has to be introduced, which classifies the persons in accordance to their raw score. The respective model is given by

$$(2.2) \qquad E(n_A) = \mu_{\#A} \, \Pi_{j \epsilon A} \, \lambda_j$$

where $\mu_{\#A}$ depends on the scoregroup with r = #A correct responses.
Although this model is obviously overparameterised with 2k+1 parameters the intrinsic dimension of the model is only 2k, since replacement of λ_j by λ_j/c and μ_r by $\mu_r c^r$, c>0, leaves the distribution unchanged and so the overparameterisation can be avoided by assuming e.g. $\Pi_j \lambda_j=1$ or $\mu_k=1$.
By appropriately transforming the likelihood L_p for the MPM TJUR showed the correspondence to the extended random model (1.10)

%

(2.3) $$L_p(\lambda_1,\ldots,\lambda_k,\mu_0,\mu_1,\ldots,\mu_k) =$$

$$\Pi_{A\varepsilon\{1,..,k\}} \exp(-\mu_{\#A}\ \Pi_{j\varepsilon A}\lambda_j)\frac{(\mu_{\#A}\ \Pi_{j\varepsilon A}\lambda_j)^{n_A}}{n_A!} =$$

$$\frac{1}{\Pi_A n_A!}\cdot\{\exp(-\Sigma_A\mu_{\#A}\Pi_{j\varepsilon A}\lambda_j)\}\cdot\{\Pi_A(\mu_{\#A}\Pi_{j\varepsilon A}\lambda_j)^{n_A}\}\cdot$$

$$\cdot\{\frac{\Pi_{r=0}^k(\mu_r\gamma_r(\lambda_1,\ldots,\lambda_k))^{n_r}}{\Pi_{r=0}^k(\mu_r\gamma_r(\lambda_1,\ldots,\lambda_k))^{n_r}}\} =$$

$$\frac{n!}{A^{n_A}!}\ \{\exp(-\Sigma_A\mu_{\#A}\Pi_{j\varepsilon A}\lambda_j)\frac{(\Sigma_A\mu_{\#A}\Pi_{j\varepsilon A}\lambda_j)^n}{n!}\}\ \cdot$$

$$\cdot\{\Pi_{r=0}^k(\frac{\mu_r\gamma_r(\lambda_1,\ldots,\lambda_k)}{\Sigma_{r'=0}^k\mu_{r'}\gamma_{r'}(\lambda_1,\ldots,\lambda_k)})^{n_r}\}\cdot\{\frac{\Pi_A(\Pi_{j\varepsilon A}\lambda_j)^{n_A}}{\Pi_{r=0}^k\gamma_r(\lambda_1,\ldots,\lambda_k)^{n_r}}\}\ \dagger)$$

----

†Notice that

$$\Sigma_A\mu_{\#A}\Pi_{j\varepsilon A}\lambda_j = \Sigma_{r'=0}^k\mu_{r'}\gamma_{r'}(\lambda_1,\ldots,\lambda_k)$$

and

$$\Pi_A\mu_{\#A}^{\ n_A} = \Pi_{r=0}^k\mu_r^{\ n_r}$$

Defining the parameter of the Poisson distribution by $\theta = \Sigma_A \mu_{\#A} \Pi_{j\epsilon A} \lambda_j$, and the probability q_r that a given subject falls in scoregroup $r = \#A$, given the total number of observed subjects n

$$q_r = 1/\theta\{\mu_r \gamma_r(\lambda_1,..,\lambda_k)$$

the likelihood L_p for the MPM becomes

$$(2.4)\ L_p =$$

$$\{\frac{n!}{\Pi_A n_A!}\}\cdot\{e^{-\theta}\frac{\theta^n}{n!}\}\cdot\{q_0^{n_0} q_1^{n_1}\ldots q_k^{n_k}\}\cdot\{\frac{\lambda_1^{s_1}\lambda_2^{s_2}\ldots\lambda_k^{s_k}}{\Pi_{r=0}^{k}\gamma_r(\lambda_1,\ldots,\lambda_k)^{n_r}}\} =$$

$$\{\frac{n!}{\Pi_A n_A!}\}\cdot\{e^{-\theta}\frac{\theta^n}{n!}\}\cdot\{q_0^{n_0} q_1^{n_1}\ldots q_k^{n_k}\}\cdot\{\quad L_c \quad\} =$$

$$\{\frac{n!}{\Pi_A n_A!}\}\cdot\{e^{-\theta}\frac{\theta^n}{n!}\}\cdot\{\quad L_e \quad\}.$$

The experiment simulated by (2.3) can be described as Follows: Apart from the combinatorial constant, first the total number of subjects n is observed, then n_r the number of subjects with exactly r correct responses given n, and finally the distribution of response patterns within the scoregroup n_A is observed.

TJUR's main result is that the likelihood L_p for the MPM (besides the part reflecting the Poisson process of sampling and the combinatorial factor describing the fact that only counts of subjects with given response patterns are classified by the contingency table) equals the likelihood L_e for the extended random model. The ML-estimates for the item parameters in the MPM coincide with those of the conditional Rasch model, since L_p decomposes (as shown in (2.4)) as a product of L_c of the item parameters and a function of the remaining parameters.

3. CONTROL OF THE MODEL

Since a certain model restricts the space of parameters representing the whole data, the divergence of the two respective likelihood functions can be applied to testing the adequacy of the chosen model structure. Using this well-known procedure of likelihood ratio statistic for the MPM formulated above is simply to test the 2^k-dimensional model against the model which assumes separate Poisson distributions for each of the 2^k cell counts. Whence

$$(3.1) \qquad -2\ln\left(\frac{\max L_p}{\Pi_A e^{-n_A} \frac{n_A^{\,n_A}}{n_A!}}\right) = -2\ln\left(\frac{\max L_p}{\max L_f}\right),$$

where L_f denotes the likelihood for the full model, is approximately χ^2-distributed with df = 2^k-2k.

As easily can be seen, anyhow, reliability of this LR-statistic will only be given by not too small values for n and/or not too large values for k, since the average of n_A is $n/2^k$. To avoiding this problem an alternative way of goodness-of-fit-test is suggested by TJUR(1980), which coincides with the LR-statistic for the conditional Rasch model as developed by ANDERSEN(1973). The original MPM specified can be tested against a model assuming separate item parameters for each scoregroup, i.e.

$$(3.2) \qquad E(n_A) = \mu_{\#A}\, \Pi_{j \varepsilon A}\, \lambda_{j,\#A}$$

The number of parameters for (3.2) is $(k+1)+(k+1)k = (k+1)^2$. However, the intrinsic dimension is only k^2-k+2, since the scoregroups for r=0 and r=k don't leave any information for estimating the parameters. The corresponding likelihood-ratio test

$$(3.3) \qquad -2\ln\left(\frac{\max L_p}{\max L_f}\right) - \left(-2\ln\left(\frac{\max L_s}{\max L_f}\right)\right) = -2\ln\left(\frac{\max L_p}{\max L_s}\right)$$

is approximately χ^2-distributed with df = $(k^2-k+2)-2k =$ $= (k-2)(k-1)$. For further possibilities of checking the MPM see TJUR(1980).

A graphical method for assessing the adequacy of the specified model is based on the property of estimating the item parameters independently of the person parameters. Whence the estimates of the λ_i's for subsets of the sample, defined by any criterion which is assumed to be meaningful, should not be different if the model holds. This can easily be proved by simply plotting the respective estimates based on different subsets of person against each other. Under null-hypothesis the estimates should not diverge from a straight line with unit slope and zero intercept. If some deviations occur the respective items can be abandoned and the whole procedure of fitting and testing the modified model might be redone iteratively to find a set of items assumable homogeneous to a latent dimension represented by the remaining items.

4. USING THE "GLIM"-PACKAGE FOR NUMERICAL SOLUTIONS

The present chapter is thought to exhibit the necessary features to use GLIM for obtaining numerical solutions for the parameters of the models discussed. Whence in addition to a brief outline of the concept of Generalised Linear Models (GLMs) emphazising the case of Poisson error the respective GLIM-directives and -macros will be presented and illustrated by an example.

4.1 GENERALISED LINEAR MODELS

The formulation of a general structure of linear approaches provides the possibility of building up various models in the context of GLMs, introduced by NELDER and WEDDERBURN(1972).

It seems appreciable therefore to formulate the connections between the RM and the MPM (as suggested by TJUR,1980) with respect to the GLM approach partly to show the relation to other linear models partly to enable the use of the GLIM-package for parameter estimation.

The global structure of GLMs is given by

$$(4.1) \qquad y = \mu + e$$

where the y's are the observations, μ represents the systematic and e the stochastic part of the model. Density of y is defined by any distribution of the exponental family, i.e.

$$(4.2) \qquad p(y) = \exp\{(y\theta - b(\theta))/a(\phi) + c(y,\phi)$$

θ is the canonical parameter and ϕ the scale parameter.

Since, as well-known

$$E(\frac{\partial L}{\partial \theta}) = 0$$

and

$$E(\frac{\partial^2 L}{\partial^2 \theta}) + E(\{\frac{\partial L}{\partial \theta}\}^2) = 0$$

where L is the log likelihood, mean and variance are given by

(4.3) $E(y) = \mu = b'(\theta)$

and

(4.4) $var(y) = \sigma = b''(\theta) \cdot a(\phi)$

A specification of the model structure (4.1) is defined by the distributional assumptions of the error component and by formulating a function of the linear predictor η, where

$$E(y) = \mu = f(\eta) = f(X\beta).$$

X represents the design matrix (covariates) of the parameters β to estimate. The linking function $f(\eta)$ relates the linear predictor $\eta = X\beta$ to the expected value μ of the observations y and can be any monotone and differentiable function.

Anyhow, each distribution has a special link (also called canonical link) so that $\theta = \eta$, which leads to desirable statistical properties (as sufficiency) and computational coincidence of the GAUSS-NEWTON-method and an iterative-weighted-least-squares-(IWLS)-algorithm as used by the GLIM-package. (For further details see NELDER,WEDDERBURN,1972 or GLIM-MANUAL, Release 3,1978)

Now furmulating the MPM of Section 2. (and so the corresponding extended randomised version of the RM) in a "GLM-terminology" the following specifications have to be made. Under (4.3) and (4.4) in the case of Poisson error (4.2) becomes

$$(4.6) \qquad p(y) = e^{-\mu} \cdot \frac{\mu^y}{y!} = \exp\{\theta y - e^{\theta} + (-\ln y!)\}$$

The canonical link is given by

$$(4.7) \qquad \eta = \ln(\mu) \quad \text{or} \quad \mu = e^{\eta}.$$

For checking the adequacy of the chosen model a discrepancy measure for this model c against the full model f (as many parameters as observations) is defined by the scaled deviance S(c,f)

$$(4.8) \qquad S(c,f) = -2\ln(L_c/L_f) = D(c,f)/\phi$$

where

$$S(c,f) = 2\{\Sigma_{i=1}^{n} \frac{y_i(\tilde{\theta}_i - \hat{\theta}_i) + b(\hat{\theta}_i) - b(\tilde{\theta}_i)}{a_i(\phi)}\}$$

D(c,f) is defined as the deviance of the current model c in relation to the full model f, where ϕ represents the scale parameter, $\hat{\tilde{\theta}}_i$ the MLE of θ_i under the current and $\hat{\theta}_i$ under the full model respectively. Since D(c,f) contains no unknowns and can be computed from the data and the MLEs directly this measure in the case of Poisson error is

$$(4.9) \qquad -2\{\Sigma(y \ln(y/\hat{\mu}) - (y-\hat{\mu}),$$

which is the LR-statistic (3.1) and equals to the G^2-test suggested by BISHOP(1969).

The alternative LR-test (3.3) which corresponds to ANDERSEN's proposal for checking the RM cannot be computed directly but its given by the difference of the deviance D(c,f) of the fit of the original MPM and the deviance D(s,f) of the extended MPM assuming separate item parameters for each scoregroup.

4.2 GLIM-DIRECTIVES AND -MACROS FOR ESTIMATING THE PARAMETERS OF THE RASCH-MODEL AND THE MULTIPLICATIVE POISSON MODEL

First of all it should be noticed that the use of GLIM is restricted to cases where only a moderate number of items, $k \leq 10$, is to be analysed. This restriction is due to space limitations. Appendix B. provides information of needed space by using the macros and directives presented in this section with respect to different samplesize (number of items and subjects). Additionally the "$ENVIRONMENT"-directive might be used for getting further informations about the local installation of GLIM.

The input to GLIM for estimating the parameters of the MPM contains two parts. One is the input of the data and the formulation of the model wanted. Since this part is only dependent on the special requirements of the respective model and data it has to be specified by the user. The second part consists of macros for estimating the parameters which are given in Appendix A. They should be added to the actual GLIM-session before any other directives are used.

In the following these two parts are discussed in more detail beginning with the part concerning the special input. An example (taken from FISCHER,1974,p.519ff.) is used partly to illustrate the input section and partly to show the coincidence of numerical solutions of a MPM by usage of GLIM and a RM, obtained by means of a widely used Fortran program (also given by FISCHER,ib.).

Suppose a sample of $n = 10$ subjects who have responded to $k = 4$ items of an ability test. The observations might be

Subjects:	Items: 1	2	3	4
1	0	1	1	0
2	1	0	1	1
3	1	1	0	0
4	1	1	1	0
5	1	0	0	0
6	0	1	1	0
7	0	0	1	0
8	0	0	1	0
9	0	1	1	1
10	0	1	0	1

where a correct answer is coded by 1, else 0.

The MPM describing the data is according to (4.1) and (4.7), in a reparameterised form neglecting the error component, given as

(4.10)

$$
\begin{bmatrix} y_{n_1} \\ y_{n_2} \\ \vdots \\ \\ \\ \\ y_{n_7} \\ \\ \\ \\ \\ \\ \\ \\ \vdots \\ y_{n_{16}} \end{bmatrix}
=
\begin{bmatrix} 0 \\ 0 \\ 2 \\ 0 \\ 0 \\ 1 \\ 2 \\ 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}
= \mathrm{EXP} \left[\underbrace{\begin{matrix}
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0 \\
1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\
1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\
1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\
1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1
\end{bmatrix} \\
GM \; I_1 \; I_2 \; I_3 \; I_4 \; F_1 \; F_2 \; F_3 \; F_4
\end{matrix}}_{X}
\cdot
\begin{bmatrix} \beta_{GM} \\ \beta_{I_1} \\ \beta_{I_2} \\ \beta_{I_3} \\ \beta_{I_4} \\ \beta_{F_1} \\ \beta_{F_2} \\ \beta_{F_3} \\ \beta_{F_4} \end{bmatrix}
\right]
$$

The vector y contains the cell frequencies, i.e. the number of subjects n_i with response pattern i, $i=1,..,2^k$. The column vector GM of X describes the fact that all y_{n_i}´s are commonly predicted by the "grand mean"-effect β_{GM}. Additionally I_1 to I_4, ordered with respect to the different response patterns i, are the covariates of the item parameters, whereas the vectors F_1 to F_4 define the "marginal"-effect. Considering response pattern 0 1 1 0, we find two observations in the data to be classified accordingly. As an example the respective cell, represented in (4.10) by line 7, has been isolated to illustrate the structure of parameters for that pattern under the current model

$$y_{n_7} = \exp(\beta_{GM} + \beta_{I_2} + \beta_{I_3} + \beta_{F_2}).$$

Now building up model (4.10) using GLIM the first directives to be specified concern the number of items k and the number of observations (units) n reflecting the length of the input vector.

```
$CALC %K=4:%N=10:UNITS %N
```
†)

Since the vector of the dependent variate in case of the MPM consists of cell frequencies but in many cases the data are given in form of a nxk-matrix as in the present example, it seemed appreciable to provide the possibility of converting this matrix into the appropriate form simply by using a macro. This macro (YVAR) treats the original data matrix as a single vector, i.e. a row vector of the data matrix is treated as a scalar. Therefore a row of the input matrix should contain only 0 or 1 and no blanks between them.

----

†) All necessary directives are framed. The parts left to the user for actual specifications are underlined. Note, that the directives should be used in the presented order!

The respective declarations are

```
$DATA D $READ
```

followed by the "data-vactor":

```
0110.
1011.
.....
0101.
```

The next step is to formulate the model and the structure of the covariates. Two macros FAC and FIT, which are used by the estimation procedure, have to be specified.

Macro FAC defines the factors (covariates) and their respective levels. The identifiers used are optional but have to correspond to those specified in macro FIT. Two declarations are required. One is to assign the appropriate values to vectors. This can easily be done by using the GLIM function $CALC x=%GL(i,j); where i and j are positive integers. The resulting vector x with unit length will be initialised then by values of 1 to i in blocks of j. In the present case i will be 2 for all covariates of the item parameters, since the corresponding factors require only 2 levels in accordance to the representation of an item in the respective response pattern. Given k items the required values for j are $2^0, 2^1, \ldots, 2^{k-1}$ for the factors I_1 to I_k.
The second declaration defines the vectors I_1 to I_k to be factors, i.e."$FAC I_1 2" enables the GLIM package to treat vector I_1 as factor having two levels. Additionally the factor for the marginals has to be introduced. The number of levels required is k+1. The identifiers F has to be used, since the assignment of the appropriate values is given by macro MARG (to be presented later), which uses F as the respective identifier.
Whence for the present example the macro to be specified is

Whence the macro to be specified for the present example is

```
$MACRO FAC
$CALC I1=%GL(2,1):I2=%GL(2,2):I3=%GL(2,4):I4=%GL(2,8)
$FAC I1 2 I2 2 I3 2 I4 2 F 5
$ENDMAC
```

Macro FIT has to contain the model formula, i.e all factors connected by the operator "+". In the present case the declarations are

```
$MACRO FIT
$FIT I1+I2+I3+I4+F
$ENDMAC
```

Now all required specifications for computing the model have been given to the GLIM system except one directive which defines the accuracy of the person parameters calculated iteratively. This is given by declaring

```
$CALC %E=0.00001
```

To start the procedure for estimating the parameters the directive

```
$USE EST
```

has to be used. Results for graphically checking the model can be obtained by

```
$USE GR
```

A summarisation of the directives used for computing the present example is given in Appendix A.

As mentioned above the macros EST and GR have to be used for getting the results of parameter estimation and the graphical test of the model.

The general structure of macro EST is given in Table 1.

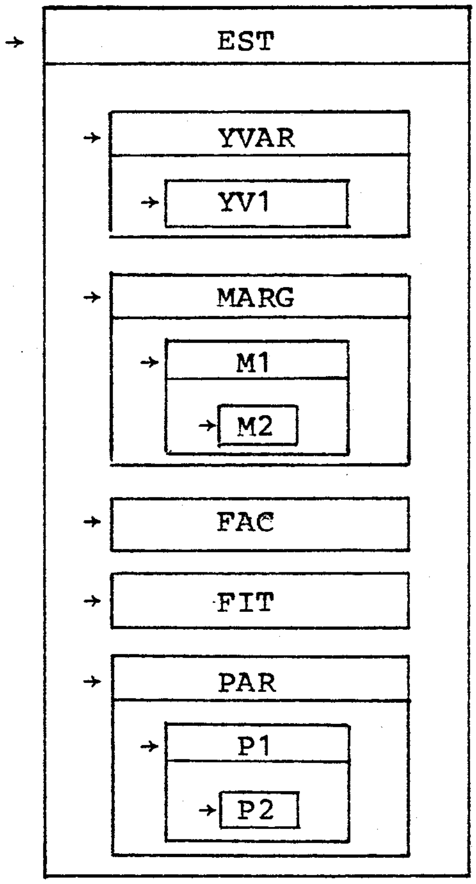

Table 1 ("→" is to read as "use of macro ...")

Macro YVAR specifies and defines the dependent variate y by transforming the original (nxk)-data matrix, treated as a single vector with length n, into the vector y with length 2^k. Macro MARG calculates the values for factor F describing the effect of the person's marginals in accordance to the ordering of the cell frequencies by macro YVAR. Usage of macros FAC and FIT provides the estimation of the model. Macro PAR extracts the item parameters and calculates the person parameters (scoregroup parameters) of the RM iteratively by solving

$$r_v = \Sigma_i \exp(\xi_v+\lambda_i)/(1+\exp(\xi_v+\lambda_i)) \qquad r_v = 1,..k-1$$

given the extracted λ_i's.
Additionally the output section is specified by macro PAR.

The second main macro is GR, providing the results of the parameter estimation for different subsamples according to RASCH's proposal for checking the model as described in Section 3. The partitioning of the sample is defined by the median of the raw marginals. Two different sets of item parameters are estimated accordingly and plotted against each other. The structure of macro GR is given in Table 2.

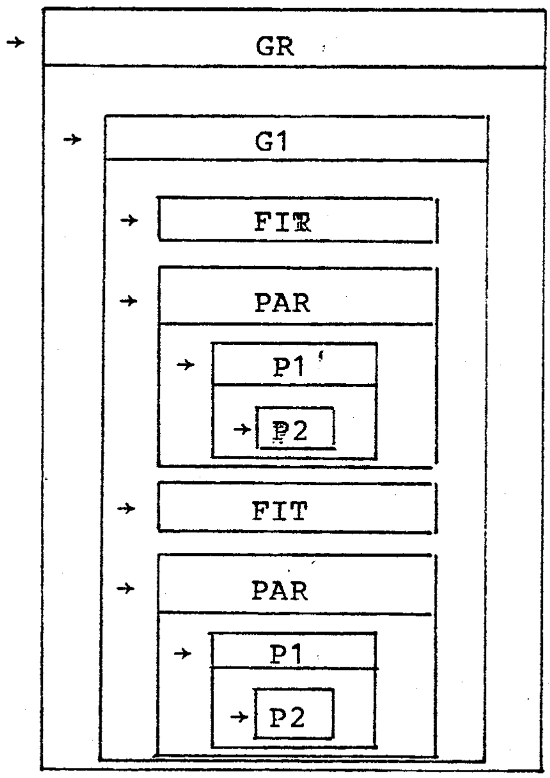

Table 2

Now having specified all required directives and macros the following output for the example presented results:

PARAMETERS FOR ALL SUBJECTS
****

CYCLE	SCALED DEVIANCE	DF
9	8.592	8

PARAMETERS FOR ITEMS

1	-.3701
2	.3701
3	.7600
4	-.7601

PARAMETERS FOR SUBJECTS

1	-1.187
2	.7775-004
3	1.187

PARAMETERS FOR SUBSAMPLES:
****

NUMBER OF SUBJECTS WITH LOW SCORE: 7.000
NUMBER OF SUBJECTS WITH HIGH SCORE: 3.000

PARAMETERS FOR SUBJECTS WITH LOW SCORE
****

CYCLE	SCALED DEVIANCE	DF
10	4.295	8

PARAMETERS FOR ITEMS

1	-.6585
2	1.041
3	.2762
4	-.6585

PARAMETERS FOR SUBJECTS

1	-1.225
2	-.1151-001
3	1.221

PARAMETERS FOR SUBJECTS WITH HIGH SCORE
****

CYCLE	SCALED DEVIANCE	DF
12	.1983-003	8

PARAMETERS FOR ITEMS

1	5.135
2	-5.481
3	5.828
4	-5.481

PARAMETERS FOR SUBJECTS

1	-5.481
2	.2774-001
3	5.481

```
...............GRAPHICAL TEST OF THE MODEL:  HIGH VS.LOW SCORE
 7.20      *
 6.40      *
 5.60      *                        H
 4.80      *       H
 4.00      *
 3.20      *
 2.40      *
 1.60      *
 .800      *
 .440-006  *                                      L
-.800      *                        L
-1.60      *       2
-2.40      *
-3.20      *
-4.00      *
-4.80      *
-5.60      *       H
-6.40      *                                      H
-7.20      *
 .........*.........*.........*.........*.........*.........*
        -1.00     -.500      .000      .500     1.000      1.50
```

ITEMPARAMETERS FOR SCOREGROUP

	LOW	HIGH
1	-.6585	5.135
2	1.041	-5.481
3	.2762	5.828
4	-.6585	-5.481

REFERENCES

ANDERSEN, E.B.: A goodness of fit test for the Rasch model. Psychometrika 38, 123-140 (1973)

ANDERSEN, E.B.: Discrete statistical models with social science applications. North-Holland P.C., Amsterdam 1980

FISCHER, G.H.: Einführung in die Theorie psychologischer Tests. Huber, Bern 1974

FISCHER, G.H.: On the existence and uniqueness of maximum-likelihood estimates in the Rasch model. Psychometrika, 46, 59-77 (1981)

NELDER, J.A., R.W.M. WEDDERBURN: Generalized linear models. Journ. Roy. Stat. Soc., A, 135, 370-384 (1972)

RASCH, G.: Probabilistic models for some intelligence and attainment tests. Copenhagen, Paedogogiske Institute, 1960

TJUR, T.: A connection between Rasch's item analysis model and a multiplicative Poisson model. Copenhagen, Institute of Math. Stat., 1980

APPENDIX A

Computional aspects:

	10 objects	100 objects	200 objects	<p36>
4 items	3319	3510	3760	11
5 items	3473	3649	3899	12
6 items	3799	4024	4274	13
7 items	4384	4609	4859	16
8 items	5577	5802	6052	20
9 items	8060	8285	8535	30
10items	13229	13454	13704	50

The table above gives values of used space for various amounts of data.

Option <p36>...... is an additional core-request parameter, an integer giving the number of blocks of 256 GLIM words . If <p36>is dummy, MAXDAT= 1000 otherwise MAXDAT= 1000 +<p36> * 256

An approximation for calculating the required data space:

$$\text{DATASPACE (DS)} = 3500 + (2^{k-1} (9+k)) + 5N/2$$

k= number of items
N= number of objects

The value for <p36> to specify is then given by (DS-1000)/256.

APPENDIX B: Macros and example

```
$MACRO EST
$USE YVAR
$USE MARG
$USE FAC
$PRINT :::::'PARAMETERS FOR ALL SUBJECTS'
$PRINT '***************************':
$YVAR Y
$ERR P
$CYCLE 50 0
$USE FIT
$USE PAR
$ENDMAC
```

```
$MACRO YVAR
$CA %X=%K:I=0:%M=2**%K:MM=0$
$WHILE %X YV1
$VAR %M Y
$CA Y=0:I=I+1:Y(I)=Y(I)+1$
$ENDMAC
$MACRO YV1
$CA %U=%K-%X:%Z=10**(%X-1):%W=2**%U:E=%TR(D/%Z)
$CA MM=MM+%EQ(E,1):I=I+%IF(%EQ(E,1),%W,0):D=D-E*%Z:%X=%X-1$
$ENDMAC
```

```
$MACRO MARG
$UNITS %M$
$CA F=0:%X=0:%Y=1:%L=-%K
$WHILE %L M1
$CA F=F+1$
$ENDMAC
$MACRO M1
$CA %J=2**(%L+%K)
$WHILE %J M2
$CA %X=0:%L=%L+1$
$ENDMAC
$MACRO M2
$CA %X=%X+1:%Y=%Y+1:F(%Y)=F(%X)+1:%J=%J-1$
$ENDMAC
```

```
$MACRO GR
$SORT R MM
$CA %M=%N/2+1:C=%LE(F,R(%M))$
$USE G1$
$ENDMAC
$MACRO G1
$CA L=Y*%EQ(C,0):%C=%CU(L)
$PRINT ::
$YVAR L
$PRINT :::::'PARAMETERS FOR SUBSAMPLES:'
$PRINT '**************************'::
$PRINT 'NUMBER OF SUBJECTS WITH LOW SCORE: '%C
$CA LL=Y*%EQ(C,1):%C=%CU(LL)
$PRINT 'NUMBER OF SUBJECTS WITH HIGH SCORE:'%C
$PRINT ::::'PARAMETERS FOR SUBJECTS WITH LOW SCORE: '
$PRINT '****************************************'::
$USE FIT
$VAR %K LOW HIGH
$USE PAR
$CA LOW=EPS
$PRINT ::
$YVAR LL
$PRINT ::::'PARAMETERS FOR SUBJECTS WITH HIGH SCORE: '
$PRINT '*****************************************'::
$USE FIT
$USE PAR
$CA HIGH=EPS
$PRINT :::::
$PRINT '................GRAPHICAL TEST OF THE MODEL:  HIGH VS.LOW SCORE'
$PL LOW HIGH LOW
$PRINT::'ITEMPARAMETERS FOR SCOREGROUP'
$PRINT'             LOW         HIGH':
$LOO LOW HIGH$
$ENDMAC


$MACRO PAR
$EXTRACT %PE
$VAR %PL SUF:%K EPS
$CA SUF=%GL(%PL,1):SUF=SUF-1:SUF=%IF(%LE(SUF,%K),SUF,0):EPS(SUF)=%PE$
$CA %U=%K-1:%X=1$
$VAR %U XI$
$CA XI=1$
$WHILE %U P1
$PRINT 'PARAMETERS FOR ITEMS':
$LOO EPS
$PRINT ::'PARAMETERS FOR SUBJECTS':
$LOO XI$
$ENDMAC
$MACRO P1
$CA %D=1$
$WHILE %D P2
$CA %U=%U-1:%X=1$
$ENDMAC
$MACRO P2
$CA S=%EXP(XI(%U)-EPS)/(1+%EXP(XI(%U)-EPS)):%S=%CU(S):%X=%U-%S+XI(%U)$
$CA %T=%SQRT((XI(%U)-%X)**2):%D=%GT(%T,%E):XI(%U)=%X$
$ENDMAC
```

```
$CA %K=4:%N=10$UNITS %N$DATA D$READ
0110
1011
1100
1110
1000
0110
0010
0010
0111
0101
$CA %E=.00001
$MACRO FIT
$FIT I1+I2+I3+I4+F$
$ENDMAC
$MACRO FAC
$CA I1=%GL(2,1):I2=%GL(2,2):I3=%GL(2,4):I4=%GL(2,8)
$FAC I1 2 I2 2 I3 2 I4 2 F 5
$ENDMAC
$USE EST
$USE GR
```